JOURNAL OF BACTERIOLOGY, June 2008, p. 4139–4146 Vol. 190, No. 12
0021-9193/08/$08.00+0 doi:10.1128/JB.01853-07
Copyright © 2008, American Society for Microbiology. All Rights Reserved.

Complete Genome Sequence of the Soil Actinomycete *Kocuria rhizophila*▽

Hiromi Takarada, Mitsuo Sekine, Hiroki Kosugi, Yasunori Matsuo, Takatomo Fujisawa, Seiha Omata, Emi Kishi, Ai Shimizu, Naofumi Tsukatani, Satoshi Tanikawa, Nobuyuki Fujita,* and Shigeaki Harayama

NITE Genome Analysis Center, Department of Biotechnology, National Institute of Technology and Evaluation (NITE), 2-10-49 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan

Received 25 November 2007/Accepted 3 April 2008

The soil actinomycete *Kocuria rhizophila* belongs to the suborder *Micrococcineae*, a divergent bacterial group for which only a limited amount of genomic information is currently available. *K. rhizophila* is also important in industrial applications; e.g., it is commonly used as a standard quality control strain for antimicrobial susceptibility testing. Sequencing and annotation of the genome of *K. rhizophila* DC2201 (NBRC 103217) revealed a single circular chromosome (2,697,540 bp; G+C content of 71.16%) containing 2,357 predicted protein-coding genes. Most of the predicted proteins (87.7%) were orthologous to actinobacterial proteins, and the genome showed fairly good conservation of synteny with taxonomically related actinobacterial genomes. On the other hand, the genome seems to encode much smaller numbers of proteins necessary for secondary metabolism (one each of nonribosomal peptide synthetase and type III polyketide synthase), transcriptional regulation, and lateral gene transfer, reflecting the small genome size. The presence of probable metabolic pathways for the transformation of phenolic compounds generated from the decomposition of plant materials, and the presence of a large number of genes associated with membrane transport, particularly amino acid transporters and drug efflux pumps, may contribute to the organism's utilization of root exudates, as well as the tolerance to various organic compounds.

Kocuria rhizophila is a coccoid, gram-positive bacterium that belongs to the family *Micrococcaceae* in the order *Actinomycetales*. The type strain of *K. rhizophila* (DSM 11926^T) was isolated from the rhizosphere of narrowleaf cattail (*Typha angustifolia*) (22). The genus *Kocuria* was created from the genus *Micrococcus* on the basis of the phylogenetic and chemotaxonomic dissection of the genus *Micrococcus* (48). *K. rosea*, *K. varians*, *K. kristinae*, *K. palustris*, *K. rhizophila*, *K. polaris*, *K. marina*, *K. himachalensis*, and *K. aegyptia* are the current validly described species (48, 22, 41, 21, 28, 25). Members of the genus *Kocuria* were isolated from a wide variety of natural sources, including mammalian skin, soil, the rhizosphere, fermented foods, clinical specimens, freshwater, and marine sediments. This is rather surprising considering its relatively small genome size among actinomycetes, suggesting that each *Kocuria* species is highly adapted to respective ecological niche. The genus *Kocuria* includes several halotolerant or phenol-degrading strains. *K. kristinae*, *K. rhizophila*, and *K. marina* tolerated up to 10% NaCl in growth media (21, 22). DeRito et al. (10) showed that primary phenol degraders in soil exposed to phenol were dominated by members of the genera *Kocuria* and *Staphylococcus*.

Kocuria species, *K. rhizophila* in particular, are also important from industrial viewpoints. *K. rhizophila* ATCC 9341, formerly *Micrococcus luteus*, is designated as a quality control strain in a number of applications, including susceptibility assays for a variety of antibiotics (49). *K. rhizophila* DC2201 (NBRC 103217) was derived from IFO 12708 and characterized as a strain exhibiting tolerance to a wide variety of organic solvents (14). The small genome size, the ability to grow rapidly and at high cell density, and the robustness of the cells at various growth conditions (14) would be highly advantageous for the development of a bacterial bioconversion system that could be used under harsh conditions such as in organic solvents.

In spite of the ecological and industrial importance, no complete genome information is currently available for the bacteria in *Kocuria*/*Micrococcus* group. Recently, genomes of two strains, *Arthrobacter aurescens* TC1 (31) and *Arthrobacter* sp. strain FB24 (http://genome.jgi-psf.org/finished_microbes/art_f/art_f.home.html), of the genus *Arthrobacter*, another dominant group within the family *Micrococcaceae*, were analyzed to completion, as well as the genome of *Renibacterium salmoninarum* ATCC 33209 (http://www.genome.washington.edu/UWGC/Projects/index.cfm?PID=167), which also belongs to *Micrococcaceae*. We present here the complete genome sequence of the soil actinomycete, *K. rhizophila* DC2201.

MATERIALS AND METHODS

Sequencing, assembly, and gap closure. A DNA shotgun library with inserts of 2 to 3 kb in pUC118 vector (Takara) was constructed, as described previously (46). Plasmid clones were end sequenced by using dye terminator chemistry on an ABI Prism 3730 sequencer as described previously (46). Raw sequence data corresponding to ~10-fold coverage were assembled by using PHRED/PHARAP/CONSED software (http://www.pharap.org) (12, 13). For assembly validation, a fosmid library with inserts of 40 kb in the pCC1FOS fosmid vector was constructed by using the CopyControl Fosmid library production kit (Epi-

* Corresponding author. Mailing address: NITE Genome Analysis Center, Department of Biotechnology, National Institute of Technology and Evaluation (NITE), 2-10-49 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan. Phone: 81-3-3481-1933. Fax: 81-3-3481-8424. E-mail: fujita-nobuyuki@nite.go.jp.

▽ Published ahead of print on 11 April 2008.

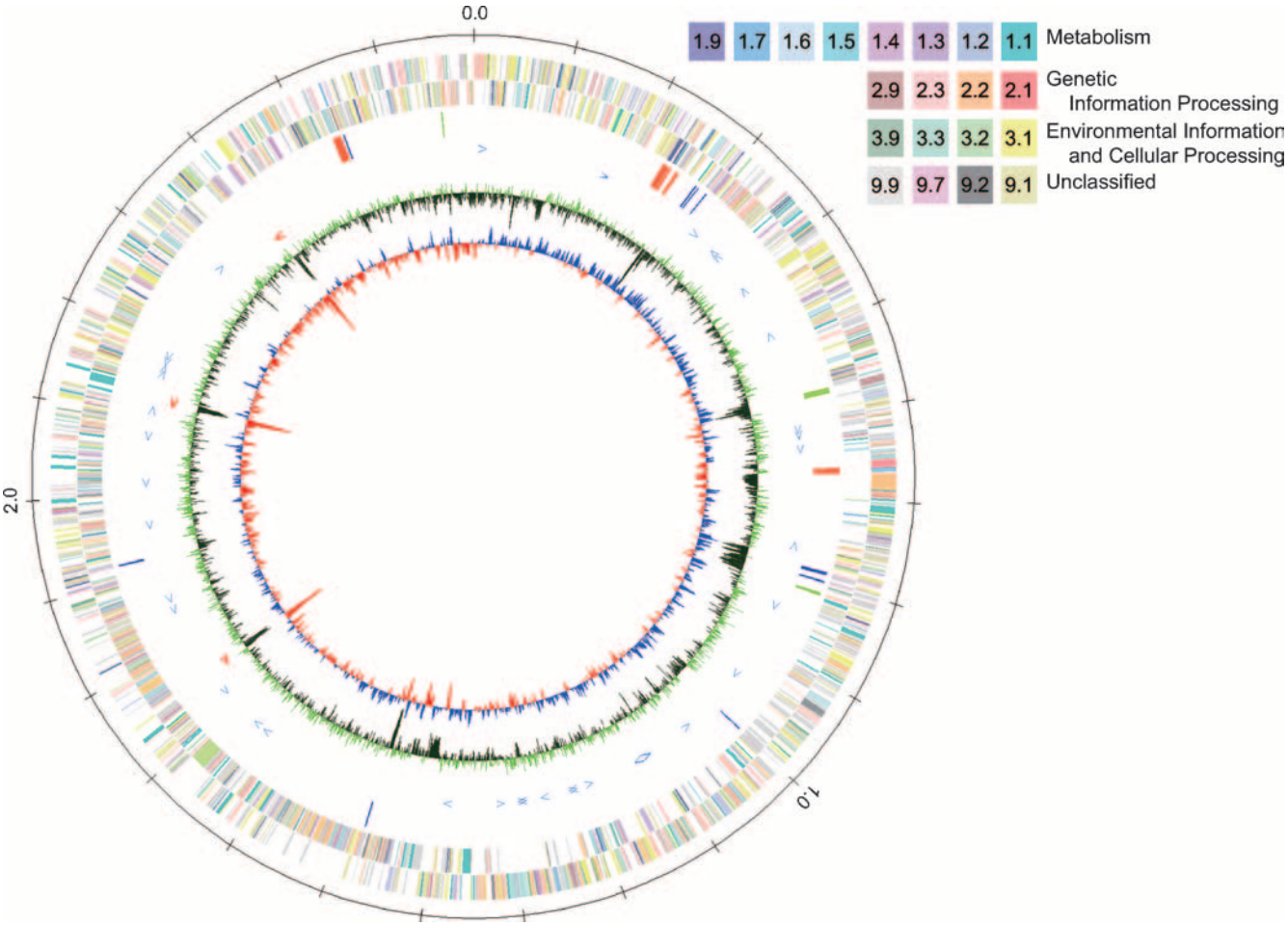

FIG. 1. Circular representation of the *K. rhizophila* chromosome. From outside to inside: circle 1, genomic position in megabases; circles 2 and 3, predicted protein coding sequences on the forward and reverse strands; circle 4, gene clusters involved in aromatic compound catabolism (red), transposon-related genes (blue), genes for restriction endonucleases (yellow green); circle 5, tRNA genes; circle 6, rRNA operons; circle 7, G+C content; circle 8, GC skew. The functional classification codes are listed in Table 2.

center). Fosmid DNA was extracted from *Escherichia coli* transformants by using a Montage BAC96 MiniPrep kit (Millipore), and end sequencing was carried out by using dye terminator chemistry on an ABI Prism 3730, as described previously (46). Fosmid end sequences were mapped onto the assembled sequence. Fosmid clones that link two contigs were selected and sequenced by primer walking to close gaps. The sequencing of difficult templates was performed by using a CUGA sequencing kit (Nippon Genetech).

Genome analysis and annotation. Putative nontranslated genes were identified by using the Rfam (15) and tRNAscan-SE (26) programs, whereas rRNA genes were identified by using the BLASTN (2) program. For the identification of protein-coding genes, the genome sequence was translated in six frames to generate potential protein products of open reading frames (ORFs) longer than 90 bp, with ATG, GTG, and TTG considered as potential initial codons. The potential protein sequences were compared to the UniProt (4) databases by using the BLASTP (2) program. Potential protein sequences that showed significant similarities to known protein sequences in the database were selected. The start sites were manually inspected and altered in comparison to the prediction obtained by GLIMMER (9, 44) and GeneHacker (55). These predicted ORFs were further evaluated by using the Frameplot program (16). The translated sequences of the predicted protein-coding genes were searched against the nonredundant UniProt database and the protein signature database, InterPro (35). The KEGG database was used for pathway reconstruction (18). Signal peptides in proteins were predicted by using SIGNALP (5), whereas transmembrane helices were predicted by using TMHMM (23).

Nucleotide sequence accession number. The complete genome sequence and annotation of *K. rhizophila* DC2201 is available at GenBank/EMBL/DDBJ under accession no. AP009152.

RESULTS AND DISCUSSION

General features of the genome. We determined the complete nucleotide sequence of the *K. rhizophila* DC2201 genome by using a whole-genome shotgun strategy, and the assembly was validated by a fosmid sequence. The genome of *K. rhizophila* DC2201 consists of a single circular chromosome of 2,697,540 bp in length with an average G+C content of 71.16% (Fig. 1 and Table 1). The general features of the genome are listed in Table 1. The chromosome encodes 46 tRNA genes, three copies of rRNA operons, and 2,357 predicted protein-coding genes. Protein functions were manually assigned based on UniProt and InterPro searches, and specific functions were predicted for 1,237 genes (52.5% of the protein-coding genes). Among the remaining predicted proteins, 225 (9.5%) were assigned to proteins belonging to specific protein families, 713 (30.3%) were assigned to hypothetical proteins (showing se-

TABLE 1. General features of *K. rhizophila* genome

Feature	Value
Length (bp)	2,697,540
G+C content (%)	71.16
Protein coding genes (no.)	2,357
Protein coding (%)	88.35
Average gene length (bp)	1,015
rRNA operons (no.)	3
tRNA genes (no.)	46
tmRNA (no.)	1

quence similarity to published proteins without known function), and 170 (7.2%) were assigned to orphans (lacking sequence similarity to published proteins). The summary of functional annotation is shown in Table 2. The taxonomic distribution of BLASTP best hits against the nonredundant UniProt database is as follows: *Actinobacteria* (2,067 genes), *Proteobacteria* (164 genes), *Firmicutes* (30 genes), and other bacteria (28 genes). The genome of *K. rhizophila* DC2201 only contains 12 transposon-related genes (Fig. 1, circle 4).

Reevaluation of taxonomic position. The suborder *Micrococcineae* is the most diverse group within the phylum *Actinobacteria* containing ecologically, morphologically and chemotaxonomically divergent bacterial species. Phylogenetic analysis based on 16S rRNA sequences is therefore not enough to resolve precise branching order or interrelationship among different subgroups. By taking advantage of the whole-genome information of *K. rhizophila*, we performed multigene phylogenetic analysis based on 122 protein genes that are conserved among various actinobacterial genomes (52). A consensus tree clearly positioned *K. rhizophila* within the same taxonomic group as *Arthrobacter* and *Renibacterium*, with *K. rhizophila* branching out at a deepest position (data not shown), although the evolutionary relationship between *Kocuria*/*Arthrobacter*/*Renibacterium* group (*Micrococcaceae*) and other *Micrococcineae* bacteria with known genome sequences, i.e., *Leifsonia xyli* and *Tropheryma whipplei*, could not be resolved with high confidence level. Consistently, dot plot analysis of the orthologous genes indicated that the genome organization of *K. rhizophila* DC2201 is most similar to that of *Arthrobacter* sp. strain FB24 (Fig. 2) despite the large difference in genome sizes.

Metabolism. From the genome sequence, *K. rhizophila* DC2201 seems to possess the enzymes required for the biosynthesis of all essential amino acids, with the exception of asparagine synthetase. In most bacteria belonging to *Actinomycetales*, the gene for histidinol-phosphate phosphatase (EC 3.1.3.15) has not been identified by similarity searches to known histidinol-phosphate genes. This enzyme catalyzes the penultimate step of histidine biosynthesis, namely, the dephosphorylation of histidinol phosphate to histidinol, which is the direct precursor of histidine. Recently, *hisN*, a novel gene that encodes an alternative form of histidinol-phosphate phosphatase, was identified in *Corynebacterium glutamicum* ATCC 13032 (33). *K. rhizophila* DC2201 is expected to catalyze the complete histidine biosynthesis pathway since it possesses the *hisN* gene homolog (KRH_09380).

Candidates for the complete sets of genes for glycolysis, the pentose phosphate pathway and the trichloroacetic acid (TCA) cycle are present, as well as those for the glyoxylate cycle for acetate catabolism and the 2-methylcitrate pathway for propionate catabolism. *K. rhizophila* DC2201 lacks the Entner-Doudoroff pathway. The annotation also suggests that this microorganism possesses the enzymes required for the biosynthesis of biotin, folate, lipoate, molybdopterin, pantothenate, pyridoxine, pyridine nucleotide, thiamine, riboflavin, thioredoxin, mycothiol (MSH), menaquinone, heme, and porphyrin, as well as those for the synthesis of all five purine and pyrimidine nucleotides; however, those required for the cobalamin biosynthesis pathway are totally absent.

TABLE 2. Summary of functional annotation

Category	No. of genes	% of total	Description
1	571	24.2	Metabolism
1.1	121	5.1	Carbohydrate metabolism
1.2	42	1.8	Energy metabolism
1.3	68	2.9	Lipid metabolism
1.4	65	2.8	Nucleotide metabolism
1.5	147	6.2	Amino acid metabolism
1.7	71	3.0	Metabolism of cofactors and vitamins
1.8	22	0.9	Biosynthesis of secondary metabolites and biodegradation of xenobiotics
1.9	34	1.4	Others
2	339	14.4	Genetic information processing
2.1	117	5.0	Transcription
2.2	119	5.0	Translation
2.3	89	3.8	Replication and repair
2.9	14	0.6	Others
3	327	13.9	Environmental information and cellular processing
3.1	226	9.6	Membrane transport
3.2	33	1.4	Signal transduction
3.3	54	2.3	Cellular processes
3.9	14	0.6	Others
9	1120	47.5	Unclassified
9.1	172	7.3	Putative function (general)[a]
9.2	53	2.2	Uncharacterized protein[a]
9.7	12	0.5	Transposon-related functions
9.8	713	30.3	Hypothetical protein[b]
9.9	170	7.2	Orphan[c]

[a] Belonging to a specific protein family.
[b] Showing similarity to published proteins with unknown function.
[c] Lacking similarity to published proteins.

Respiration. *K. rhizophila* grows under strictly aerobic conditions (22). The NADH generated during the oxidation of the carbon sources seems to be oxidized by the product of the *ndh* gene (KRH_17770), which encodes a membrane-bound, non-proton-pumping, single-subunit NADH dehydrogenase. The genome of *K. rhizophila* DC2201 lacks the *nuo* genes encoding a proton-pumping NADH dehydrogenase. However, the *qcr* gene homologs encoding menaquinol-cytochrome *c* reductase (complex III, KRH_12810-12830) and *cta* gene homologs encoding cytochrome *c* oxidase (complex IV, KRH_12800, 12850, 12860) are present. Malate:quinone oxidoreductase (KRH_10160), succinate dehydrogenases (KRH_17900-17930, 07510-07530), formate dehydrogenase (KRH_07770+07780, 07790), and glycerol-3-phosphate dehydrogenase (KRH_13520) may transfer electrons to menaquinone. Two complete yet different sets of genes presumably encoding complex II (succinate dehydrogenase) are present in the genome; gene clusters KRH_

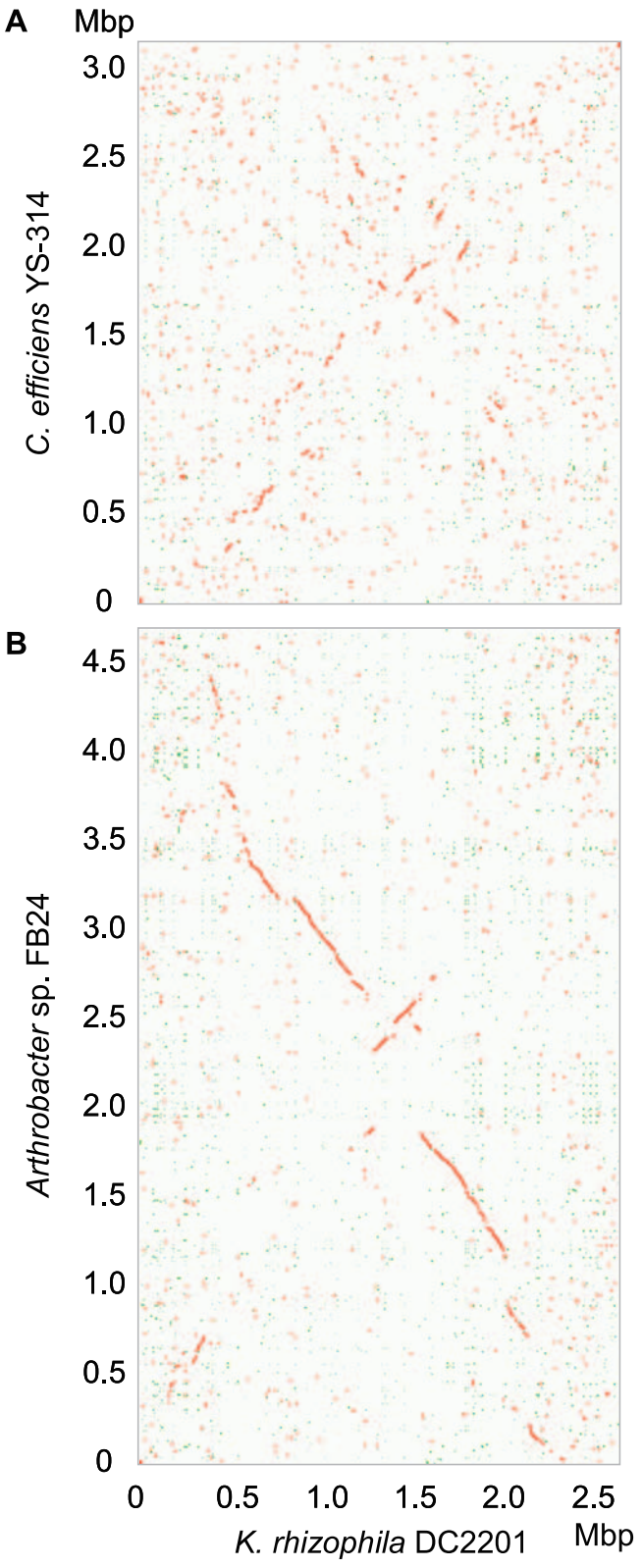

FIG. 2. Synteny between the genomes of *K. rhizophila* and *C. efficiens* (A) and *Arthrobacter* sp. strain FB24 (B). Each genome is adjusted to the *dnaA* gene as the zero point of the chromosome. Red and green dots indicate reciprocal best hits and BLASTP hits, respectively.

17900-17930 and KRH_07510-07530 are supposed to encode *E. coli*-type and low-GC-gram-positive-bacteria-type succinate dehydrogenases, respectively. Low-GC-gram-positive-bacteria-type succinate dehydrogenase from *B. subtilis* functions as a fumarate reductase in anaerobic conditions (7, 45). Two adjacent ORFs (KRH_07770+07780) collectively show 36% identity to the selenocysteine- and tungsten-containing formate dehydrogenase large subunit of *Desulfovibrio gigas*. Although *D. gigas* is a sulfate-reducing bacterium considered to be a strict anaerobe, the formate dehydrogenase activity in soluble extract was air insensitive (1). Formate dehydrogenase also allows *E. coli* to use formate as major electron donor during nitrate respiration. Consequently, *K. rhizophila* DC2201 seems to have the potential to proliferate under anaerobic conditions.

Catabolism of aromatic compounds. Genome analysis of *K. rhizophila* DC2201 reveals the presence of probable catabolic pathways for phenylacetate, protocatechuate, and homoprotocatechuate to TCA cycle intermediates (Fig. 3). Protocatechuate and homoprotocatechuate are derivatives of botanical aromatic compounds (vanillate, cafeate, farulate, *p*-cumarate, lignin, etc.) released during the decomposition of plant materials. Phenylacetate is degraded into succinyl-coenzyme A (CoA) and acetyl-CoA via β-ketoadipyl-CoA by the *paa* gene products (KRH_02100-02170, 02200, 02280-02300). *K. rhizophila* DC2201 appears to convert homoprotocatechuate to succinate by the *meta*-cleavage pathway encoded by the genes clustered in the genome (KRH_22000-22050), whereas the enzymes encoded by the *pca* genes (KRH_06040-06100) probably convert protocatechuate to succinyl-CoA and acetyl-CoA (β-ketoadipate central pathway). The gene (KRH_22060) adjacent to the probable *meta*-cleavage gene cluster may encode phenol 2-monooxygenase (EC 1.14.13.7) participating in the oxidation of phenol derivatives into *o*-diols, which are then channeled into the homoprotocatechuate catabolic pathway. The enzyme from *Trichosporon cutaneum*, a soil-living yeast, is known to possess a broad substrate specificity hydroxylating simple hydroxyl-, amino-, methyl-, and halogen-substituted phenols (17). The *K. rhizophila* DC2201 enzyme is similar (35% identity) to the eukaryotic enzyme from *T. cutaneum*, although the exact substrate specificity of the *K. rhizophila* enzyme is not known.

Secondary metabolism. The *K. rhizophila* DC2201 genome seems to encode only a limited number of secondary metabolic enzymes; a type III polyketide synthase (PKS) and a nonribosomal peptide synthetase. The organism thus possesses a much smaller number of secondary metabolic genes among the actinomycetes. The genome does not contain genes for typical bacterial PKS (type I or II). The potential type III PKS of *K. rhizophila* DC2201 encoded by KRH_07690 shares ca. 30% identity with proteins related to the plant-specific PKSs of the phytoalexin and chalcone synthase family (3, 29). A similar protein has been discovered in the balhimycin biosynthetic gene cluster in *Amycolatopsis mediterranei* (37). The potential nonribosomal peptide synthetase encoded by KRH_11450 is a protein smaller than usual nonribosomal peptide synthetases (1,382 residues), which consists of an amino acid adenylation domain, a phosphopantetheine-binding domain, and a nonribosomal peptide synthetase C-terminal domain. Candidate genes for phosphopantetheinyl transferase (KRH_11470) and thioesterase (KRH_11480) were found near the nonribosomal

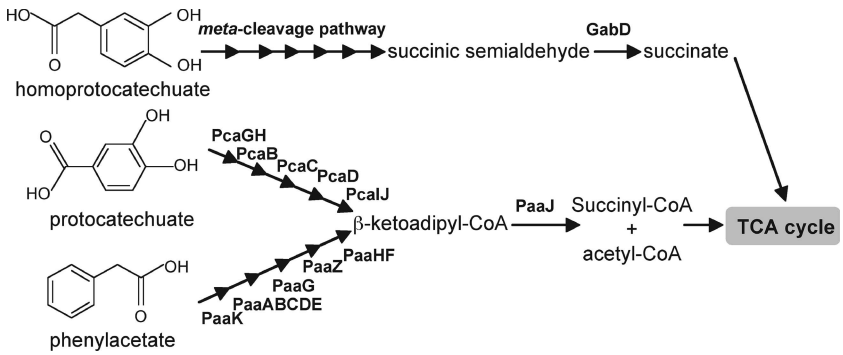

FIG. 3. Predicted biochemical steps for catabolism of aromatic compounds in *K. rhizophila* DC2201. Enzymes involved in the *meta*-cleavage pathway are homoprotocatechuate dioxygenase, 5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase, 2-hydroxyhepta-2,4-diene-1,7-dioic acid isomerase, 5-carboxymethyl-2-oxo-hex-3-ene-1,7-dioic acid decarboxylase, 2-oxo-hepta-3-ene-1,7 dioic acid hydratase, and 2,4-dihydroxy-hept-2-ene-1,7-dioic acid aldolase. The enzymes involved in the protocatechuate (Pca) and phenylacetate (Paa) catabolism are described in the text.

peptide synthetase gene locus, but the functions of these genes are not yet clear. Unlike other actinomycetes such as *Streptomyces* and *Mycobacterium* species, whose genomes typically encode more than 20 copies of cytochrome P450 monooxygenase genes (52), the *K. rhizophila* DC2201 genome contains only a single copy of the P450 gene (KRH_21570) that may be involved in fatty acid metabolism.

Information processing and modification systems. In addition to DNA polymerase I (KRH_11870) and III (KRH_14710, 00020, 02830, 13090, 18820) with multiple copies of the genes encoding the epsilon subunit of polymerase III (KRH_08570, 10050), the organism seems to possess *dnaE2* (KRH_11560), a paralog of *dnaE* (KRH_14710), and encode an error-prone DNA polymerase. DnaE2 is not a member of the Y family of error-prone DNA polymerases but is known to participate in error-prone DNA repair synthesis and contribute to the emergence of drug resistance in *Mycobacterium tuberculosis* (6). The presence of multiple copies of the major replicative DNA polymerase (DnaE) is known in some organisms, including pathogens and symbionts. The genome of *K. rhizophila* DC2201 also contains a *dinB* homolog (KRH_14930), which may encode the Y family error-prone DNA polymerase IV.

K. rhizophila DC2201 seems to contain at least one type I and two type IV restriction endonucleases. The type IV systems are composed of one or two genes encoding proteins that cleave only modified DNA, including methylated, hydroxymethylated, and glucosyl-hydroxymethylated bases (42). The restriction of methylated DNA transformation has been observed in *E. coli* (54), *Streptococcus pneumoniae* (24), *Streptomyces* strains (27), and several types of coryneform bacteria (53). The *K. rhizophila* DC2201 genome contains the genes probably encoding type IV Mrr (KRH_23230) and McrBC (KRH_07470, 07480) methyl-specific restriction systems, which may serve to control the entry and expression of foreign DNA using methylation patterns as a recognition criterion. McrBC is the only known GTP-dependent restriction enzyme (11, 38).

Stress response and tolerance. *K. rhizophila* DC2201 seems to encode fewer transcriptional regulatory proteins than other actinomycetes. These include the primary sigma factor (KRH_14360), three ECF-type sigma factors (KRH_05520, 09420, 21970), and 116 transcriptional regulators. When classified into biological role categories based on the COG database (50), the percentages of protein-coding genes in the transcription category (K) are 3.9% (*K. rhizophila* DC2201), 7.1% (*Corynebacterium glutamicum*), 5.8% (*M. tuberculosis*), and 8.0% (*Leifsonia xyli*) (Fig. 4). The genome of *K. rhizophila* DC2201 contains a moderate number of genes for two-component systems (10 complete systems), including the probable MtrA-MtrB (KRH_08800, 08810) and RegX3-SenX3 (KRH_18740, 18750) systems. The MtrA-MtrB system is involved in osmoprotection (30), and *betP* (KRH_03970), *proP* (KRH_03010), and *mscL* (KRH_06500) may be controlled by this system. Although the RegX3-SenX3 system is involved in the virulence of *M. tuberculosis* (36), the function of the system in *K. rhizophila* remains unknown.

The genes for heme-containing catalases (KRH_05290, 05500, 22630), manganese-containing superoxide dismutase (KRH_00400), peroxiredoxins (KRH_10180, 10830), and thiol peroxidase (KRH_08020) may be involved in oxidative stress tolerance. The MSH-dependent response to oxidative stress also seems to be functional in *K. rhizophila* DC2201. MSH is the dominant low-molecular-weight thiol, a reducing agent and the storage form of cysteine, which is produced by mycobacteria and other actinomycetes (51). Candidates for the genes encoding the complete MSH biosynthetic pathway (KRH_06620, 08220, 13920, 05710) and MSH reductase (KRH_15030) are present in the genome. MSH S-conjugate amidase (Mca; KRH_17690), involved in the MSH-dependent detoxification of xenobiotics such as alkylating agents, electrophiles, and antibiotics, also seems to be encoded in the genome.

Dormancy. *M. luteus* is a nonsporulating bacterium closely related to *K. rhizophila* and can persist in a dormant state after prolonged incubation in stationary phase. Dormant *M. luteus* cells lose their ability to grow, but the addition of supernatant from growing *M. luteus* cultures to dormant cells restores their ability to divide freely, thereby resuscitating the cells to normal colony-forming bacteria (34). Such dormant cells are also observed in *M. tuberculosis* and *Rhodococcus rhodochrous* (47). *M. luteus* cells secrete a resuscitation-promoting factor (Rpf), which promotes the resuscitation of dormant cells (34). Genes encoding Rpf-like proteins are widely distributed throughout

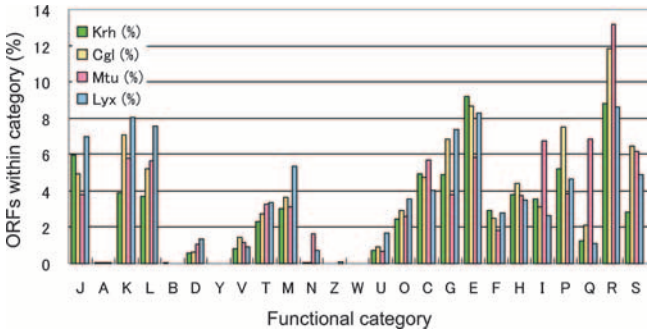

FIG. 4. Distribution of genes within the functional categories based on the COG database. The percentage (of total protein coding sequences) of genes within each category for *K. rhizophila* DC2201 (Krh), *C. glutamicum* (Cgl), *M. tuberculosis* (Mtu), and *Leifsonia xyli* (Lyx) is shown. Functional categories: J, translation, ribosomal structure, and biogenesis; A, RNA processing and modification; K, transcription; L, replication, recombination, and repair; B, chromatin structure and dynamics; D, cell cycle control, cell division, and chromosome partitioning; Y, nuclear structure; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; Z, cytoskeleton; W, extracellular structures; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, and chaperones; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, biosynthesis, transport, and catabolism of secondary metabolites; R, prediction of general function; S, function unknown.

the actinomycetes. The solution structure of an Rpf domain from *M. tuberculosis* showed homology to lysozymes, suggesting that oligosaccharide cleavage is the signal for revival from dormancy (8). The gene encoding a putative Rpf precursor (KRH_16710) is also present in the genome of *K. rhizophila* DC2201. The Rpf polypeptide (228 residues) is composed of three domains: N-terminal signal peptide, central Rpf domain, and C-terminal LysM peptideglycan-binding domain. We assume that *K. rhizophila* DC2201 can promote the resuscitation of dormant cells persisting under stressful or nutrient-limited conditions. *K. rhizophila* DC2201 genome also possess possible homologs (KRH_19960, 19970) of *hipBA* encoding a toxin-antitoxin module known to be involved in the formation of persister cells and multidrug resistance in *E. coli* (20). These mechanisms may confer tolerance to environmental stresses in *K. rhizophila*.

Membrane transport. More than 200 genes (9.6%) in *K. rhizophila* DC2201 seem to be involved in membrane transport, such as P-P-bond-hydrolysis-driven transporters, electrochemical potential-driven transporters, channels/pores, and phosphotransferase systems (Table 3).

The genome of *K. rhizophila* DC2201 contains a surprisingly large number of amino acid-polyamine-organocation (APC) family transporters, particularly considering its relatively small genome size (Table 4). APC family transporters function as solute:cation symporters and solute:solute antiporters (43). We predict that the APC transporter family in *K. rhizophila* DC2201 includes L-asparagine permeases (KRH_19430, 21050), D-serine/D-alanine/glycine transporter (KRH_04050), lysine-specific permease (KRH_02030), proline-specific permease (KRH_22910), aromatic amino acid transport proteins (KRH_01990, 20200), gamma-aminobutyrate permease (KRH_19500), ethanolamine permease (KRH_00930), and unknown amino acid transporters (KRH_00240, 00880, 00900, 14170, 17350, 22350).

TABLE 3. Transporters in the genome of *K. rhizophila* DC2201

Transporter type	No. of genes	% of total
Transport proteins	226	9.6
ATP dependent	86	3.6
ATP-binding cassette (ABC) superfamily	83	3.5
Secondary transporters	115	4.9
MFS	41	1.7
APC family	15	0.6
Ion channels	4	0.2
Group translocators	5	0.2
Protein secretion	14	0.6
Unclassified	2	0.08

The genome also contains 13 proteins possibly involved in the multidrug resistance efflux system: 11 proteins are members of the major facilitator superfamily (MFS), and 2 proteins (encoded by KRH_22090 and KRH_22100) may constitute two-component multidrug efflux pumps, which are members of the small multidrug resistance family. These efflux systems, which catalyze the active extrusion of many structurally and functionally unrelated compounds from bacterial cytoplasm to the external medium (39), are assumed to be involved in the transport of toxic organic compounds, including a wide range of organic solvents to which the DC2201 strain is known to be tolerant. One of the MFS proteins in *K. rhizophila* (KRH_15620) shares significant homology with the AlbF multidrug efflux pump of the plant pathogen *Xanthomonas albilineans*. The AlbF system exports the potent bacterial and plant toxin albicidin produced by *X. albilineans*. The presence of the AlbF homolog in *K. rhizophila*, as in the case of *Leifsonia xyli* (32), may suggest the association of this bacterium with plant environments where the bacterium share the same niche with plant pathogens. The resistance-nodulation-cell division family exporters, which are important for the efflux of toxic organic compounds in gram-negative bacteria (40), were not found in the genome of *K. rhizophila* DC2201.

K. rhizophila DC2201 seems to possess several protein export and secretion systems, including Sec pathway genes (KRH_05830, 06350, 08850, 12130, 13610-13620), twin-argin-

TABLE 4. Comparison of encoded APC family transporters among the genomes of *K. rhizophila* DC2201 and other selected bacteria

Organism	Genome size[a] (Mbp)	No. of protein-coding genes[a]	APC family transporters[b]	
			No.	% of total
Streptomyces avermitilis MA-4680	9.12	7,673	27	0.35
Streptomyces coelicolor A3(2)	9.05	8,154	16	0.20
Kocuria rhizophila DC2201	2.70	2,357	15	0.64
Propionibacterium acnes KPA171202	2.56	2,297	14	0.61
Nocardia farcinica IFM10152	6.29	5,936	11	0.19
Mycobacterium bovis AF2122/97	4.35	3,920	9	0.23
Mycobacterium tuberculosis H37Rv	4.41	3,989	9	0.23
Corynebacterium efficiens YS-314	3.15	2,950	5	0.17
Corynebacterium glutamicum ATCC 13032	3.30	2,993	4	0.13
Leifsonia xyli CTCB07	2.58	2,030	4	0.20
Corynebacterium diphtheriae NCTC 13129	2.49	2,272	3	0.13
Escherichia coli K-12-MG1655	4.64	4,243	22	0.52
Pseudomonas putida KT2440	6.18	5,350	21	0.39
Bacillus subtilis 168	4.21	4,105	18	0.44

[a] From the NCBI Entrez Genome Project (www.ncbi.nlm.nih.gov/genomes/).
[b] From TransportDB (www.membranetransport.org).

ine translocation (Tat) pathway genes (KRH_08370, 13790-13800), and signal recognition particle-dependent pathway genes (KRH_10560, 10590).

Survival strategies encoded in the small genome. *K. rhizophila* DC2201 has one of the smallest genome among the actinomycetes. Most proteins (87.7%) showed high similarity to those of other actinomycetes, and the primary metabolic pathways seem to be very similar to those of other actinomycetes. The genome contains smaller numbers of genes encoding transposon-related proteins, transcriptional regulators, and proteins involved in the biosynthesis of secondary metabolites. The genome may thus represent some minimal requirement for the actinomycetes.

On the other hand, some of the *K. rhizophila* specific features of the genome suggest strong association of this microorganism with plant environments, the rhizosphere, and possible survival strategies under various stress conditions. For example, the genome contains probable biodegradation pathways of phenylacetate, protocatechuate, and homoprotocatechuate, aromatic compounds that may be formed during the decomposition of plant materials, to TCA cycle intermediates. The presence of the homologs of AlbF multidrug efflux pump of plant pathogens and the plant-related type III PKS also suggests the interaction with plant environments. A number of culture-based and culture-independent studies of microbial diversity revealed that rhizosphere communities significantly differ from bulk soil communities (19). Of particular interest in this context is the large number of uncharacterized genes encoding proteins of unknown function (9.5% of coding genes) and hypothetical proteins (37.2%), some of which may encode novel proteins involved in the formation of the microbial community in the rhizosphere. The tolerance to various stress conditions, including salts and toxic hydrocarbons and alcohols, may be attributed, at least in part, to the presence of an osmoprotection system controlled by the MtrA-MtrB two-component system and the presence of a number of multidrug resistance efflux systems. A longer survival could be also achieved by the formation of dormant cells.

ACKNOWLEDGMENT

This study was supported by the New Energy and Industrial Technology Development Organization (Japan).

REFERENCES

1. **Almendra, M. J., C. D. Brondino, O. Gavel, A. S. Pereira, P. Tavares, S. Bursakov, R. Duarte, J. Caldeira, J. J. Moura, and I. Moura.** 1999. Purification and characterization of a tungsten-containing formate dehydrogenase from *Desulfovibrio gigas*. Biochemistry **38:**16366–16372.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **An, C., Y. Ichinose, T. Yamada, Y. Tanaka, T. Shiraishi, and H. Oku.** 1993. Organization of the genes encoding chalcone synthase in *Pisum sativum*. Plant Mol. Biol. **21:**789–803.
4. **Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. S. Yeh.** 2004. UniProt: the universal protein knowledgebase. Nucleic Acids Res. **32:**D115–D119.
5. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783–795.
6. **Boshoff, H. I., M. B. Reed, C. E. Barry III, and V. Mizrahi.** 2003. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. Cell **113:**183–193.
7. **Cecchini, G., E. Maklashina, V. Yankovskaya, T. M. Iverson, and S. Iwata.** 2003. Variation in proton donor/acceptor pathways in succinate:quinone oxidoreductases. FEBS Lett. **545:**31–38.
8. **Cohen-Gonsaud, M., P. Barthe, C. Bagneris, B. Henderson, J. Ward, C. Roumestand, and N. H. Keep.** 2005. The structure of a resuscitation-promoting factor domain from *Mycobacterium tuberculosis* shows homology to lysozymes. Nat. Struct. Mol. Biol. **12:**270–273.
9. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
10. **DeRito, C. M., G. M. Pumphrey, and E. L. Madsen.** 2005. Use of field-based stable isotope probing to identify adapted populations and track carbon flow through a phenol-degrading soil microbial community. Appl. Environ. Microbiol. **71:**7858–7865.
11. **Dila, D., E. Sutherland, L. Moran, B. Slatko, and E. A. Raleigh.** 1990. Genetic and sequence organization of the *mcrBC* locus of *Escherichia coli* K-12. J. Bacteriol. **172:**4888–4900.
12. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
13. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
14. **Fujita, K., T. Hagishita, S. Kurita, Y. Kawakura, Y. Kobayashi, A. Matsuyama, and H. Iwahashi.** 2006. The cell structural properties of *Kocuria rhizophila* for aliphatic alcohol exposure. Enzyme Microb. Technol. **39:**511–518.
15. **Griffiths-Jones, S., A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy.** 2003. Rfam: an RNA family database. Nucleic Acids Res. **31:**439–441.

16. **Ishikawa, J., and K. Hotta.** 1999. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. FEMS Microbiol. Lett. **174:**251–253.
17. **Kalin, M., H. Y. Neujahr, R. N. Weissmahr, T. Sejlitz, R. Johl, A. Fiechter, and J. Reiser.** 1992. Phenol hydroxylase from *Trichosporon cutaneum*: gene cloning, sequence analysis, and functional expression in *Escherichia coli*. J. Bacteriol. **174:**7112–7120.
18. **Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori.** 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. **32:**D277–D280.
19. **Kent, A. D., and E. W. Triplett.** 2002. Microbial communities and their interactions in soil and rhizosphere ecosystems. Annu. Rev. Microbiol. **56:**211–236.
20. **Keren, I., D. Shah, A. Spoering, N. Kaldalu, and K. Lewis.** 2004. Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. J. Bacteriol. **186:**8172–8180.
21. **Kim, S. B., O. I. Nedashkovskaya, V. V. Mikhailov, S. K. Han, K. O. Kim, M. S. Rhee, and K. S. Bae.** 2004. *Kocuria marina* sp. nov., a novel actinobacterium isolated from marine sediment. Int. J. Syst. Evol. Microbiol. **54:**1617–1620.
22. **Kovacs, G., J. Burghardt, S. Pradella, P. Schumann, E. Stackebrandt, and K. Marialigeti.** 1999. *Kocuria palustris* sp. nov. and *Kocuria rhizophila* sp. nov., isolated from the rhizoplane of the narrow-leaved cattail (*Typha angustifolia*). Int. J. Syst. Bacteriol. **49:**167–173.
23. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
24. **Lacks, S., and B. Greenberg.** 1977. Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. J. Mol. Biol. **114:**153–168.
25. **Li, W. J., Y. Q. Zhang, P. Schumann, H. H. Chen, W. N. Hozzein, X. P. Tian, L. H. Xu, and C. L. Jiang.** 2006. *Kocuria aegyptia* sp. nov., a novel actinobacterium isolated from a saline, alkaline desert soil in Egypt. Int. J. Syst. Evol. Microbiol. **56:**733–737.
26. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
27. **MacNeil, D. J.** 1988. Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. J. Bacteriol. **170:**5607–5612.
28. **Mayilraj, S., R. M. Kroppenstedt, K. Suresh, and H. S. Saini.** 2006. *Kocuria himachalensis* sp. nov., an actinobacterium isolated from the Indian Himalayas. Int. J. Syst. Evol. Microbiol. **56:**1971–1975.
29. **Melchior, F., and H. Kindl.** 1990. Grapevine stilbene synthase cDNA only slightly differing from chalcone synthase cDNA is expressed in *Escherichia coli* into a catalytically active enzyme. FEBS Lett. **268:**17–20.
30. **Moker, N., M. Brocker, S. Schaffer, R. Kramer, S. Morbach, and M. Bott.** 2004. Deletion of the genes encoding the MtrA-MtrB two-component system of *Corynebacterium glutamicum* has a strong influence on cell morphology, antibiotics susceptibility and expression of genes involved in osmoprotection. Mol. Microbiol. **54:**420–438.
31. **Mongodin, E. F., N. Shapir, S. C. Daugherty, R. T. DeBoy, J. B. Emerson, A. Shvartzbeyn, D. Radune, J. Vamathevan, F. Riggs, V. Grinberg, H. Khouri, L. P. Wachett, K. E. Nelson, and M. J. Sadowsky.** 2006. Secrets of soil servival revealed by the genome sequence of *Arthrobacter aurescens* TC1. PLoS Genet. **2:**e214.
32. **Monteiro-Vitorello, C. B., L. E. Camargo, M. A. Van Sluys, J. P. Kitajima, D. Truffi, A. M. do Amaral, R. Harakava, J. C. de Oliveira, D. Wood, M. C. de Oliveira, C. Miyaki, M. A. Takita, A. C. da Silva, L. R. Furlan, D. M. Carraro, G. Camarotte, N. F. Almeida, Jr., H. Carrer, L. L. Coutinho, H. A. El-Dorry, M. I. Ferro, P. R. Gagliardi, E. Giglioti, M. H. Goldman, G. H. Goldman, E. T. Kimura, E. S. Ferro, E. E. Kuramae, E. G. Lemos, M. V. Lemos, S. M. Mauro, M. A. Machado, C. L. Marino, C. F. Menck, L. R. Nunes, R. C. Oliveira, G. G. Pereira, W. Siqueira, A. A. de Souza, S. M. Tsai, A. S. Zanca, A. J. Simpson, S. M. Brumbley, and J. C. Setúbal.** 2004. The genome sequence of the gram-positive sugarcane pathogen *Leifsonia xyli* subsp. *xyli*. Mol. Plant-Microbe Interact **17:**827–836.
33. **Mormann, S., A. Lomker, C. Ruckert, L. Gaigalat, A. Tauch, A. Puhler, and J. Kalinowski.** 2006. Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. BMC Genomics **7:**205.
34. **Mukamolova, G. V., O. A. Turapov, K. Kazarian, M. Telkov, A. S. Kaprelyants, D. B. Kell, and M. Young.** 2002. The *rpf* gene of *Micrococcus luteus* encodes an essential secreted growth factor. Mol. Microbiol. **46:**611–621.
35. **Mulder, N. J., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bradley, P. Bork, P. Bucher, L. Cerutti, R. Copley, E. Courcelle, U. Das, R. Durbin, W. Fleischmann, J. Gough, D. Haft, N. Harte, N. Hulo, D. Kahn, A. Kanapin, M. Krestyaninova, D. Lonsdale, R. Lopez, I. Letunic, M. Madera, J. Maslen, J. McDowall, A. Mitchell, A. N. Nikolskaya, S. Orchard, M. Pagni, C. P. Ponting, E. Quevillon, J. Selengut, C. J. Sigrist, V. Silventoinen, D. J. Studholme, R. Vaughan, and C. H. Wu.** 2005. InterPro, progress and status in 2005. Nucleic Acids Res. **33:**D201–205.
36. **Parish, T., D. A. Smith, G. Roberts, J. Betts, and N. G. Stoker.** 2003. The *senX3-regX3* two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology **149:**1423–1435.
37. **Pfeifer, V., G. J. Nicholson, J. Ries, J. Recktenwald, A. B. Schefer, R. M. Shawky, J. Schroder, W. Wohlleben, and S. Pelzer.** 2001. A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (*S*)-3,5-dihydroxyphenylglycine. J. Biol. Chem. **276:**38370–38377.
38. **Pieper, U., D. H. Groll, S. Wunsch, F. U. Gast, C. Speck, N. Mucke, and A. Pingoud.** 2002. The GTP-dependent restriction enzyme McrBC from *Escherichia coli* forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat. Biochemistry **41:**5245–5254.
39. **Putman, M., H. W. van Veen, and W. N. Konings.** 2000. Molecular properties of bacterial multidrug transporters. Microbiol. Mol. Biol. Rev. **64:**672–693.
40. **Ramos, J. L., E. Duque, M. T. Gallegos, P. Godoy, M. I. Ramos-Gonzalez, A. Rojas, W. Teran, and A. Segura.** 2002. Mechanisms of solvent tolerance in gram-negative bacteria. Annu. Rev. Microbiol. **56:**743–768.
41. **Reddy, G. S., J. S. Prakash, V. Prabahar, G. I. Matsumoto, E. Stackebrandt, and S. Shivaji.** 2003. *Kocuria polaris* sp. nov., an orange-pigmented psychrophilic bacterium isolated from an Antarctic cyanobacterial mat sample. Int. J. Syst. Evol. Microbiol. **53:**183–187.
42. **Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. Degtyarev, D. T. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Kruger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P. C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. Vitor, G. G. Wilson, and S. Y. Xu.** 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. **31:**1805–1812.
43. **Saier, M. H., Jr.** 2000. Families of transmembrane transporters selective for amino acids and their derivatives. Microbiology **146**(Pt. 8)**:**1775–1795.
44. **Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White.** 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. **26:**544–548.
45. **Schnorpfeil, M., I. G. Janausch, S. Biel, A. Kroger, and G. Unden.** 2001. Generation of a proton potential by succinate dehydrogenase of *Bacillus subtilis* functioning as a fumarate reductase. Eur. J. Biochem. **268:**3069–3074.
46. **Sekine, M., S. Tanikawa, S. Omata, M. Saito, T. Fujisawa, N. Tsukatani, T. Tajima, T. Sekigawa, H. Kosugi, Y. Matsuo, R. Nishiko, K. Imamura, M. Ito, H. Narita, S. Tago, N. Fujita, and S. Harayama.** 2006. Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ. Microbiol. **8:**334–346.
47. **Shleeva, M. O., K. Bagramyan, M. V. Telkov, G. V. Mukamolova, M. Young, D. B. Kell, and A. S. Kaprelyants.** 2002. Formation and resuscitation of "non-culturable" cells of *Rhodococcus rhodochrous* and *Mycobacterium tuberculosis* in prolonged stationary phase. Microbiology **148:**1581–1591.
48. **Stackebrandt, E., C. Koch, O. Gvozdiak, and P. Schumann.** 1995. Taxonomic dissection of the genus *Micrococcus*: *Kocuria gen. nov.*, *Nesterenkonia* gen. nov., *Kytococcus* gen. nov., *Dermacoccus* gen. nov., and *Micrococcus* Cohn 1872 gen. emend. Int. J. Syst. Bacteriol. **45:**682–692.
49. **Tang, J. S., and P. M. Gillevet.** 2003. Reclassification of ATCC 9341 from *Micrococcus luteus* to *Kocuria rhizophila*. Int. J. Syst. Evol. Microbiol. **53:**995–997.
50. **Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. **29:**22–28.
51. **Ung, K. S., and Y. Av-Gay.** 2006. Mycothiol-dependent mycobacterial response to oxidative stress. FEBS Lett. **580:**2712–2716.
52. **Ventura, M., C. Canchaya, A. Tauch, G. Chandra, G. F. Fitzgerald, K. F. Chater, and D. van Sinderen.** 2007. Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. Microbiol. Mol. Biol. Rev. **71:**495–548.
53. **Vertes, A. A., M. Inui, M. Kobayashi, Y. Kurusu, and H. Yukawa.** 1993. Presence of *mrr*- and *mcr*-like restriction systems in coryneform bacteria. Res. Microbiol. **144:**181–185.
54. **Waite-Rees, P. A., C. J. Keating, L. S. Moran, B. E. Slatko, L. J. Hornstra, and J. S. Benner.** 1991. Characterization and expression of the *Escherichia coli* Mrr restriction system. J. Bacteriol. **173:**5207–5219.
55. **Yada, T., and M. Hirosawa.** 1996. Gene recognition in cyanobacterium genomic sequence data using the hidden Markov model. Proc. Int. Conf. Intell. Syst. Mol. Biol. **4:**252–260.